Bayesian Epistemology

Arc Kocurek

1 What is it?

Bayesian epistemology (also called "Bayesianism") is a quantitative approach to epistemology that models rational learning and inductive reasoning using probability theory. It consists of three core tenets:

> **Bayesianism:**
>
> 1. Beliefs are not binary: they come in *degrees*. Degrees of belief are also called **credences**.
> 2. Rational credences obey the axioms of *probability*.
> 3. Rational belief update goes via conditionalization.

We will break each of these assumptions down. But first, some basic clarifications:

Question: *Didn't Hume show that induction isn't justified?*

Maybe. But let's ignore that.

Question: *Is Bayesianism a descriptive theory about how people actually reason or a normative theory about how they should reason?*

It depends. Most Bayesians treat it as a normative theory, but some also treat is as a descriptive theory.

Question*: Is it realistic to assume that people have such fine-grained degrees of belief? Can we really say so-and-so believes it's raining to degree 0.7325?*

Probably not. These tenets of Bayesian epistemology are more like idealizations than absolute laws.

Question: *I don't like numbers.*

Not a question, but don't worry: you do not need a deep mathematical grasp of probability theory. You just need (a) a basic understanding of how to reason with probability, and (b) familiarity with the notation.

2 Credences

An agent's **credence** in some proposition P is, very roughly, a measure of how likely that agent thinks P is to be true. In other words, it is a measure of the strength of that agent's belief in P.

> **Example 2.1.** Consider the following propositions.
>
> (i) The campus bells will ring at some point today.
>
> (ii) The number of particles in the universe right now is even.
>
> (iii) I am a lizard.
>
> My guess is that the following is more-or-less accurate:
>
> - Your credence in (i) is fairly high. You are not *absolutely certain* that it is true; for example, maybe the bells are broken. But you think it is quite likely they will ring and would even be willing to bet money on it.
> - Your credence in (ii) is middling. You are *indifferent* on the question of whether (ii) is true. That is, you don't really have a strong opinion about (ii) either way.
> - Your credence in (iii) is pretty low. I mean, I *could* be a lizard; maybe I am a very literate lizard wearing a sophisticated disguise. But you probably wouldn't bet money on it.

Question: How can we measure the credence someone has in a proposition? Maybe we both have high credence in (i) above. But how do we know whether your credence in (i) is higher or lower than mine?

Answer: One response (due to Frank Ramsey) is to relate credences to betting behaviors. Thus, we might try to measure an agent's credence in a proposition by asking them which of two bets they would rather take.

Example 2.2. Consider the proposition that it will reach 70°F tomorrow. Let's say that I am willing to accept a bet on which I earn $1 if it reaches 70°F tomorrow and lose $9 if it does not. You, on the other hand, are not willing to accept such a bet. Then it's reasonable to infer that my credence in the proposition that it will reach 70°F tomorrow is higher than your credence in that proposition. (We might even say that I am at least 90% confident that it will reach 70°F tomorrow.)

3 Probability

Bayesianism assumes that credences behave more-or-less like probabilities. Where A is some proposition, we write "$\Pr(A)$" for the credence an agent has in A.

Credences all obey the following three axioms:

Axiom 1. For any proposition A, $0 \leqslant \Pr(A) \leqslant 1$.

This is more-or-less a matter of convention. We could in principle have chosen different ranges, but this would make the math unnecessarily complicated.

Axiom 2. If A is a necessary truth, then $\Pr(A) = 1$.

This is an idealization about rational agents. We assume that a rational agent is a *perfect logician*, meaning they are completely confident in every logical truth. It is an open question what the best way of relaxing this constraint is for mere mortals like us.

Axiom 3. If A and B are mutually exclusive propositions (i.e., A and B cannot both be true), then writing "$A \vee B$" for the proposition that either A or B (or both) is true, $\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

This axiom is also known as *finite additivity*. It is the key axiom that connects the credences of multiple propositions. We can state the axiom equivalently as follows: if $A_1, \ldots, A_n$ are mutually exclusive propositions, then $\Pr(A_1 \vee \cdots \vee A_n) = \Pr(A_1) + \cdots + \Pr(A_n)$.

From these three axioms, the following obtain for any propositions A and B:

(a) $\Pr(\neg A) = 1 - \Pr(A)$.

Proof: Since A and $\neg A$ are mutually exclusive, $\Pr(A \vee \neg A) = \Pr(A) + \Pr(\neg A)$. But since $A \vee \neg A$ is a necessary truth, $\Pr(A \vee \neg A) = 1$. Hence, $\Pr(A) + \Pr(\neg A) = 1$, and so $\Pr(\neg A) = 1 - \Pr(A)$.

(b) If A is impossible, then $\Pr(A) = 0$.

Proof: If A is impossible, then $\neg A$ is a necessary truth. Hence, $\Pr(\neg A) = 1$. Using (a), we get $\Pr(A) = 1 - 1 = 0$.

(c) If A necessarily entails B, then $\Pr(A) \leqslant \Pr(B)$.

Proof: If A necessarily entails B, then A and $\neg B$ are mutually exclusive. Hence, $\Pr(A \vee \neg B) = \Pr(A) + \Pr(\neg B) = \Pr(A) + (1 - \Pr(B))$. Since any probability must be less than or equal to 1, that means $\Pr(A \vee \neg B) \leqslant 1$. So $\Pr(A) + (1 - \Pr(B)) \leqslant 1$. Subtracting 1 from both sides, we get $\Pr(A) - \Pr(B) \leqslant 0$, i.e., $\Pr(A) \leqslant \Pr(B)$.

Exercise: Verify (d)–(g) (you may use any of the results above).

(d) If A is necessarily equivalent to B, then $\Pr(A) = \Pr(B)$.

(e) $\Pr(A) = \Pr(A \,\&\, B) + \Pr(A \,\&\, \neg B)$.

(f) $\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \,\&\, B)$.

(g) $\Pr(A \,\&\, B) \leqslant \Pr(A)$.

You might think that $\Pr(A \,\&\, B) = \Pr(A) \cdot \Pr(B)$. But actually, this is only true is certain special circumstances. We will return to this matter momentarily.

4 Conditionalization

Suppose an agent is not completely certain whether or not A is true. If they learn that A is true, what will their new credences look like? That is, what is their credence in some proposition B *given* A?

Let us write "$\Pr(B \mid A)$" to denote an agent's credence in B given A. Then the Bayesian says that an agent's credence in B once they learn for certain that A is true is determined by the following equation:

$$\Pr(B \mid A) = \frac{\Pr(A \,\&\, B)}{\Pr(A)}.$$

In other words, your credence in B when you learn A is just the proportion of your original credence in A that B takes up. Note that this definition assumes that $\Pr(A) \neq 0$ (you can't divide by 0). If $\Pr(A) = 0$, then $\Pr(B \mid A)$ is undefined.

Bayesian epistemology stipulates that belief update goes via conditionalization. That is, if $\Pr_{\text{old}}$ is an agent's credence function prior to learning that A is true and $\Pr_{\text{new}}$ is their credence function after learning that A is true, then Bayesianism postulates that:

$$\Pr_{\text{new}}(B) = \Pr_{\text{old}}(B \mid A).$$

Aside *(Jeffrey Conditionalization):* The standard conditionalization rule only applies when you learn *for certain* that A is true. But what happens if you simply become more confident that A is true without being absolutely certain?

Richard Jeffrey famously answered this question by stating a more general update rule. The idea is simple: one should adjust one's credences in a proposition B proportionally to how much one's credence in A shifts. This is captured by the following equation:

$$\Pr_{\text{new}}(B) = \Pr_{\text{old}}(B \mid A) \cdot \Pr_{\text{new}}(A) + \Pr_{\text{old}}(B \mid \neg A) \cdot \Pr_{\text{new}}(\neg A).$$

Notice that this general equation implies the simple conditionalization rule when $\Pr_{\text{new}}(A) = 1$. By contrast, if your credence in A doesn't change ($\Pr_{\text{new}}(A) = \Pr_{\text{old}}(A)$), then neither does your credence in B ($\Pr_{\text{new}}(B) = \Pr_{\text{old}}(B)$).

By a simple rearrangement, we have the following:

$$\Pr(A \,\&\, B) = \Pr(B \mid A) \cdot \Pr(A).$$

This is the most general equation governing credences in conjunctions. In particular, we do not generally have $\Pr(A \,\&\, B) = \Pr(A) \cdot \Pr(B)$, since generally $\Pr(B \mid A) \neq \Pr(B)$.

Example 4.1. Suppose I roll a pair of fair dice. You learn that the first die landed on 6. What is your credence in the proposition that the second die landed on 6? Intuitively, it's still 1/6: the two rolls are independent of each other.

Now suppose the dice are loaded. They are both biased in favor of landing on the same number, but you don't know which number that is. Now if you learn the first die landed on 6, what is your credence in the proposition that the second landed on 6? Intuitively, it's higher than 1/6: the fact that the first landed on 6 makes it more likely that they are both loaded towards 6.

If $\Pr(B \mid A) = \Pr(B)$, that means that the likelihood of B does not depend on whether A is true. We will say that B is **probabilistically independent** of A if $\Pr(B \mid A) = \Pr(B)$.

Fact 4.2 (***Equivalent Reformulations of Independence*****).**
The following are equivalent for any A and B:

(a) $\Pr(B \mid A) = \Pr(B)$.

(b) $\Pr(A \mid B) = \Pr(A)$.

(c) $\Pr(A \,\&\, B) = \Pr(A) \cdot \Pr(B)$.

(d) $\Pr(B \mid A) = \Pr(B \mid \neg A)$.

Exercise: Verify that (a)–(d) are equivalent.

The most useful formulation of probabilistic independence is (c): A and B are independent if and only if $\Pr(A \,\&\, B) = \Pr(A) \cdot \Pr(B)$. It's worth emphasizing again: in general, $\Pr(A \,\&\, B) \neq \Pr(A) \cdot \Pr(B)$. We cannot even say whether $\Pr(A \,\&\, B) \leqslant \Pr(A) \cdot \Pr(B)$ or *vice versa*.

5 Bayes's Theorem

Theorem 5.1 (***Bayes's Theorem***)**.**

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}.$$

Proof: Using the definition of conditional probability, we get:

$$\Pr(A \,\&\, B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B \,\&\, A) = \Pr(A \mid B) \cdot \Pr(B).$$

But $A \,\&\, B$ and $B \,\&\, A$ are logically equivalent, so $\Pr(A \,\&\, B) = \Pr(B \,\&\, A)$. Therefore:

$$\Pr(B \mid A) \cdot \Pr(A) = \Pr(A \mid B) \cdot \Pr(B).$$

Rearranging, we get:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}.$$

Bayes's theorem is a quite useful result for the purposes of studying confirmation. Replacing 'A' with 'H' and replacing 'B' with 'E', the theorem states:

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \cdot \Pr(H)}{\Pr(E)}.$$

This is useful is because often the probability on the righthand side of the equation are each individually easier to determine directly than the probability on the lefthand side. The lefthand side tells us how much an observation of E would confirm hypothesis H. On the righthand side, we have $\Pr(E \mid H)$, $\Pr(H)$ and $\Pr(E)$. That is, we just need to ask how likely we think E and H are now, and then determine how likely E is given the hypothesis H, which is usually just a matter of fleshing out the empirical consequences of a theory (that might still be hard, of course; but determining $\Pr(H \mid E)$ directly is usually harder).

6 Dutch Book

Bayesianism is committed to the following three assumptions:

1. Rational agents have *degrees* of belief, or credences.
2. Rational credences obey the probability axioms.
3. Rational belief update goes via conditionalization.

Assumption 1 seems fairly plausible. But why should we accept assumptions 2 or 3? That is, why should our beliefs obey the probability axioms, and why should we update our beliefs via conditionalization?

An influential answer (from Ramsey and de Finetti) utilizes *Dutch book arguments*. A **Dutch book** is a collection of bets such that if one accepts all of these bets, then one is guaranteed to lose money no matter what the outcome is. Intuitively, if an agent accepts a Dutch book, they are irrational. Given this assumption, one can prove the following: an agent will never accept a Dutch book if and only if their credences obey the axioms of probability.

For instance, suppose an agent violates Axiom 2 (that $\Pr(A) = 1$ if A is necessary). Say their credence the necessary truth A is 0.8. Then they will accept a bet on which they lose \$5 if A is true and earn \$1,000,000 if A is false. In that case, they will be *guaranteed* to lose money because A is necessarily true, i.e., $\neg A$ is impossible.

Here is an example of a Dutch book argument for Axiom 3.

Example 6.1. Let's suppose that for some mutually exclusive A and B, Dumbo's credences are: $\Pr(A \vee B) = \Pr(A) = \Pr(B) = 1/4$. Then Dumbo will take these bets if you offer them:

(1) If A, then you pay Dumbo \$8; if $\neg A$, Dumbo pays you \$2.

(2) If B, then you pay Dumbo \$8; if $\neg B$, Dumbo pays you \$2.

(3) If $A \vee B$, then Dumbo pays you \$8; if $\neg(A \vee B)$, you pay Dumbo \$3.

If he takes all three of these bets, then Dumbo owes you money no matter what happens. *Exercise:* Find a Dutch book against Dumbo if instead $\Pr(A \vee B) = 3/4$.